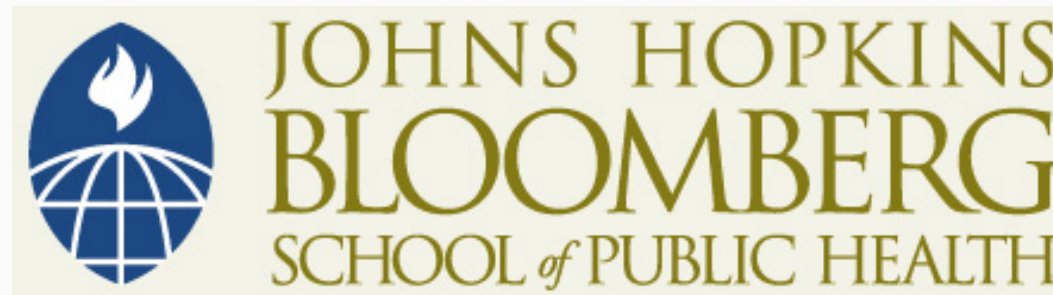


This work is licensed under a [Creative Commons Attribution-NonCommercial-ShareAlike License](https://creativecommons.org/licenses/by-nc-sa/4.0/). Your use of this material constitutes acceptance of that license and the conditions of use of materials on this site.



Copyright 2009, The Johns Hopkins University and John McGready. All rights reserved. Use of these materials permitted only in accordance with license rights granted. Materials provided “AS IS”; no representations or warranties provided. User assumes all responsibility for use, and all liability related thereto, and must independently review all materials for accuracy and efficacy. May contain materials owned by others. User is responsible for obtaining permissions for use from third parties as needed.



JOHNS HOPKINS
BLOOMBERG
SCHOOL *of* PUBLIC HEALTH

Comparing Means among Two (or More) Independent Populations

John McGready
Johns Hopkins University

Lecture Topics

- CIs for mean difference between two independent populations
- Two sample t-test
- Non-parametric alternative, Mann Whitney (FYI, optional)
- Comparing means amongst more than two independent populations:
ANOVA



JOHNS HOPKINS
BLOOMBERG
SCHOOL *of* PUBLIC HEALTH

Section A

Two Sample t-test: The Resulting Confidence Interval

Comparing Two Independent Groups

- “A Low Carbohydrate as Compared with a Low Fat Diet in Severe Obesity”*
 - 132 severely obese subjects randomized to one of two diet groups
 - Subjects followed for a six month period
- At the end of study period
 - “Subjects on the low-carbohydrate diet lost more weight than those on a low fat diet (95% confidence interval for the difference in weight loss between groups, -1.6 to -6.2 kg; $p < .01$)”

Comparing Two Independent Groups: Diet Types Study

- Scientific question
 - Is weight change associated with diet type?

	Diet Group	
	Low-Carb	Low-Fat
Number of subjects (n)	64	68
Mean weight change (kg) Post-diet less pre-diet	-5.7	-1.8
Standard deviation of weight changes (kg)	8.6	3.9

Diet Type and Weight Change

- 95% CIs for weight change by diet group
- Low Carb: $-5.7 \pm 2 \times \frac{8.6}{\sqrt{64}} \rightarrow -5.7 \pm 2 \times 1.08 \approx (-7.8 \text{ kg}, -3.5 \text{ kg})$
- Low Fat: $-1.8 \pm 2 \times \frac{3.9}{\sqrt{68}} \rightarrow -1.8 \pm 2 \times .47 \approx (-2.7 \text{ kg}, -0.9 \text{ kg})$

Comparing Two Independent Groups: Diet Types Study

- In statistical terms, is there a non-zero difference in the average weight change for the subjects on the low-fat diet as compared to subjects on the low-carbohydrate diet?
 - 95% CIs for each diet group mean weight change do not overlap, but how do you quantify for the difference?
- The comparison of interest is not “paired”
 - There are different subjects in each diet group
- For each subject a change in weight (after diet—before weight) was computed
 - However, the authors compared the changes in weight between two independent groups!

Comparing Two Independent Groups

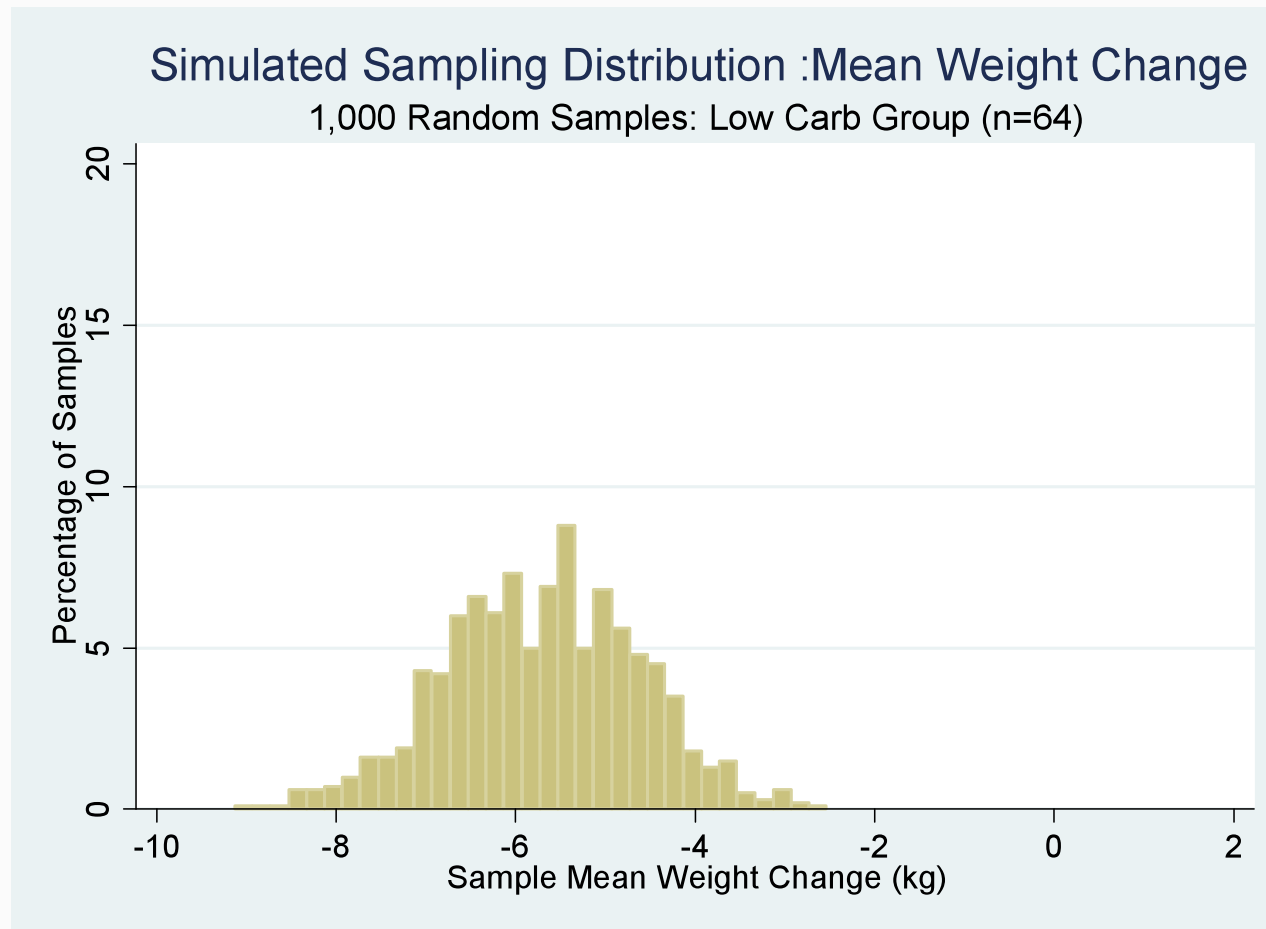
- How do we calculate
 - Confidence interval for difference?
 - p-value to determine if the difference in two groups is “significant?”
- Since we have large samples (both greater than 60) we know the sampling distributions of the sample means in both groups are approximately normal
- It turns out the difference of quantities, which are (approximately) normally distributed, are also normally distributed

Sampling Distribution: Difference in Sample Means

- So, the big news is . . .
 - The sampling distribution of the difference of two sample means, each based on large samples, approximates a normal distribution
 - This sampling distribution is centered at the true mean difference, $\mu_1 - \mu_2$

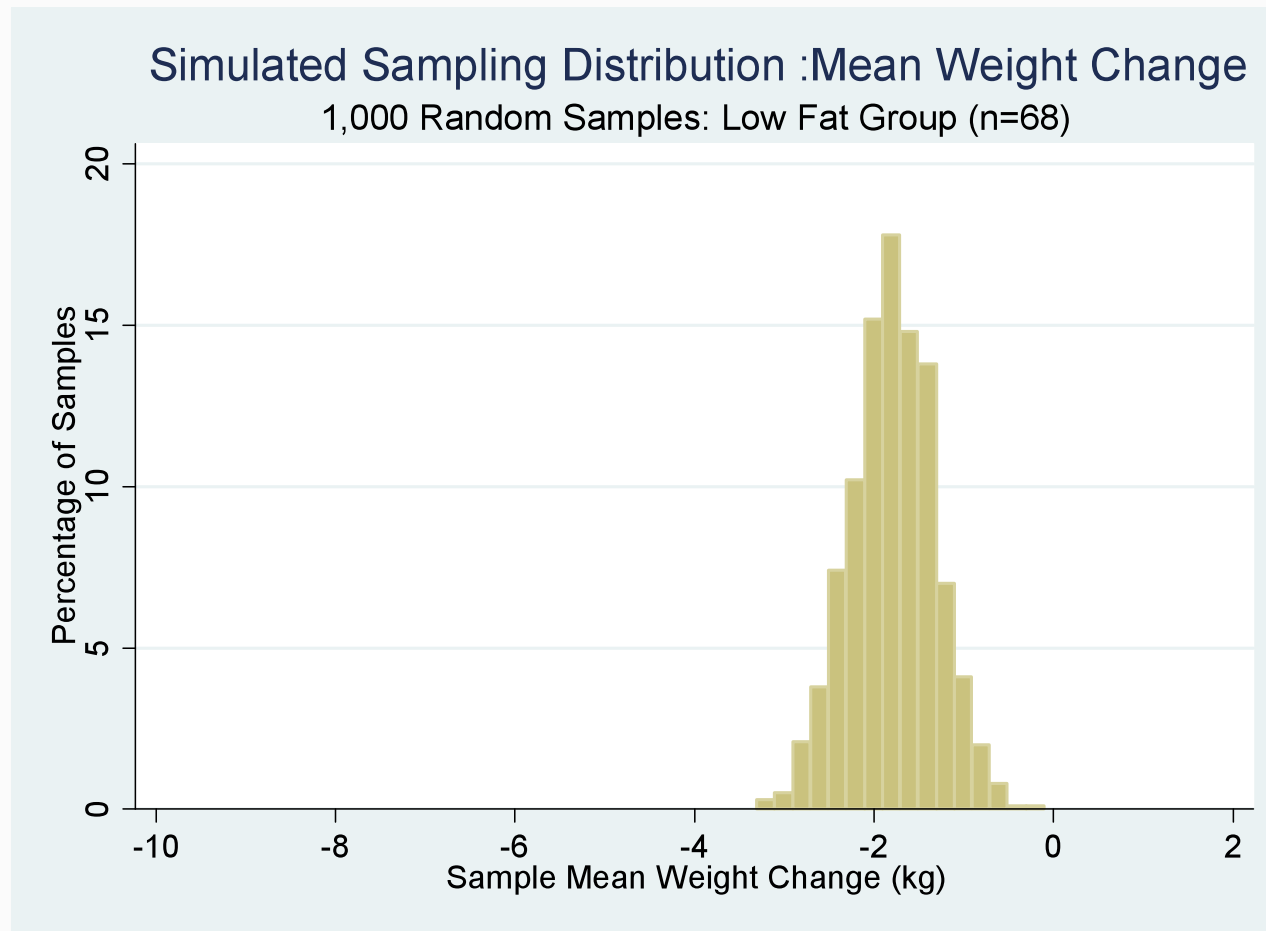
Simulated Sampling Dist'n of Sample Mean Weight Loss

- Simulated sampling distribution of sample mean weight change: low carbohydrate diet group



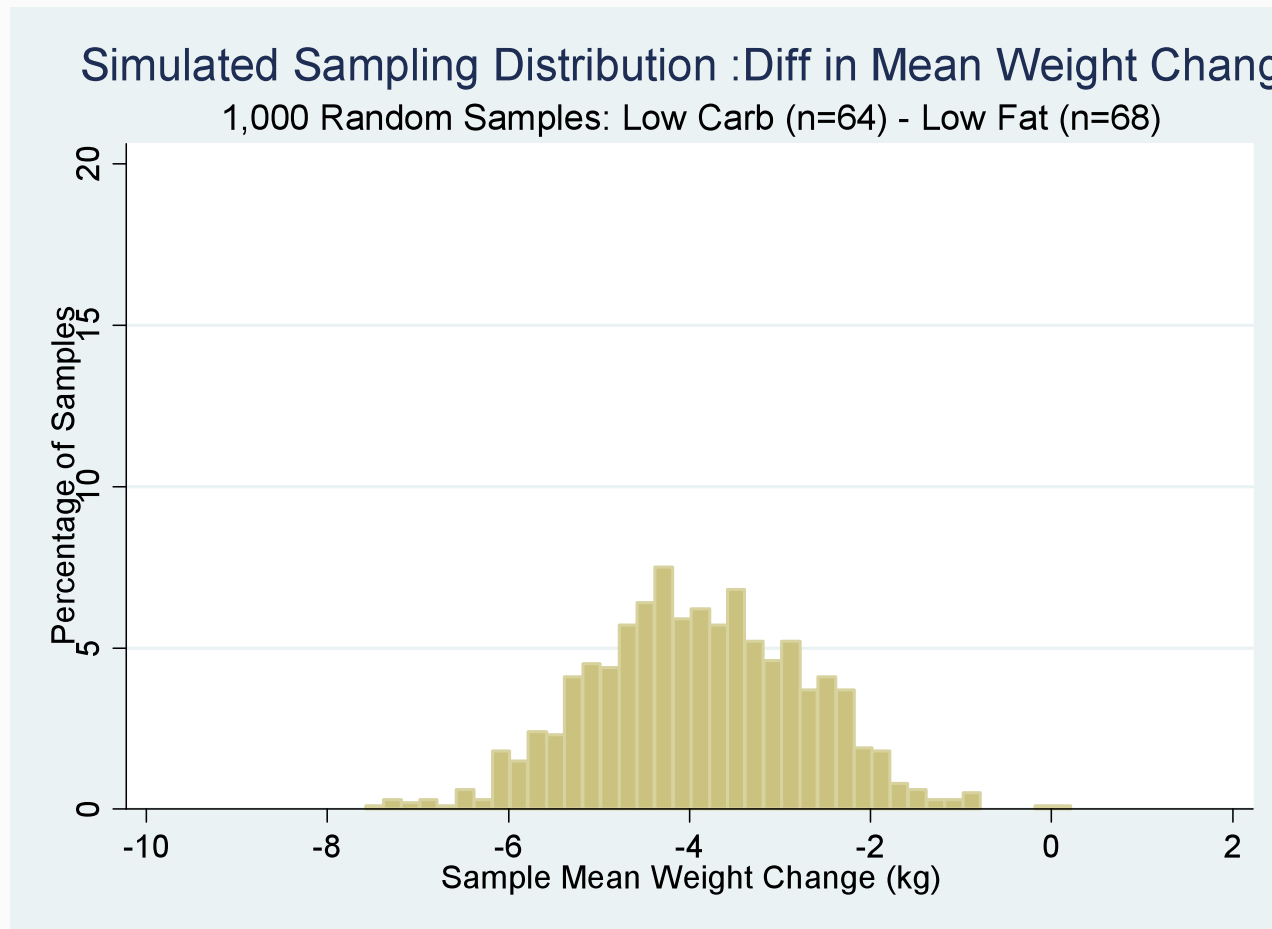
Simulated Sampling Dist'n of Sample Mean Weight Loss

- Simulated sampling distribution of sample mean weight change: low fat diet group



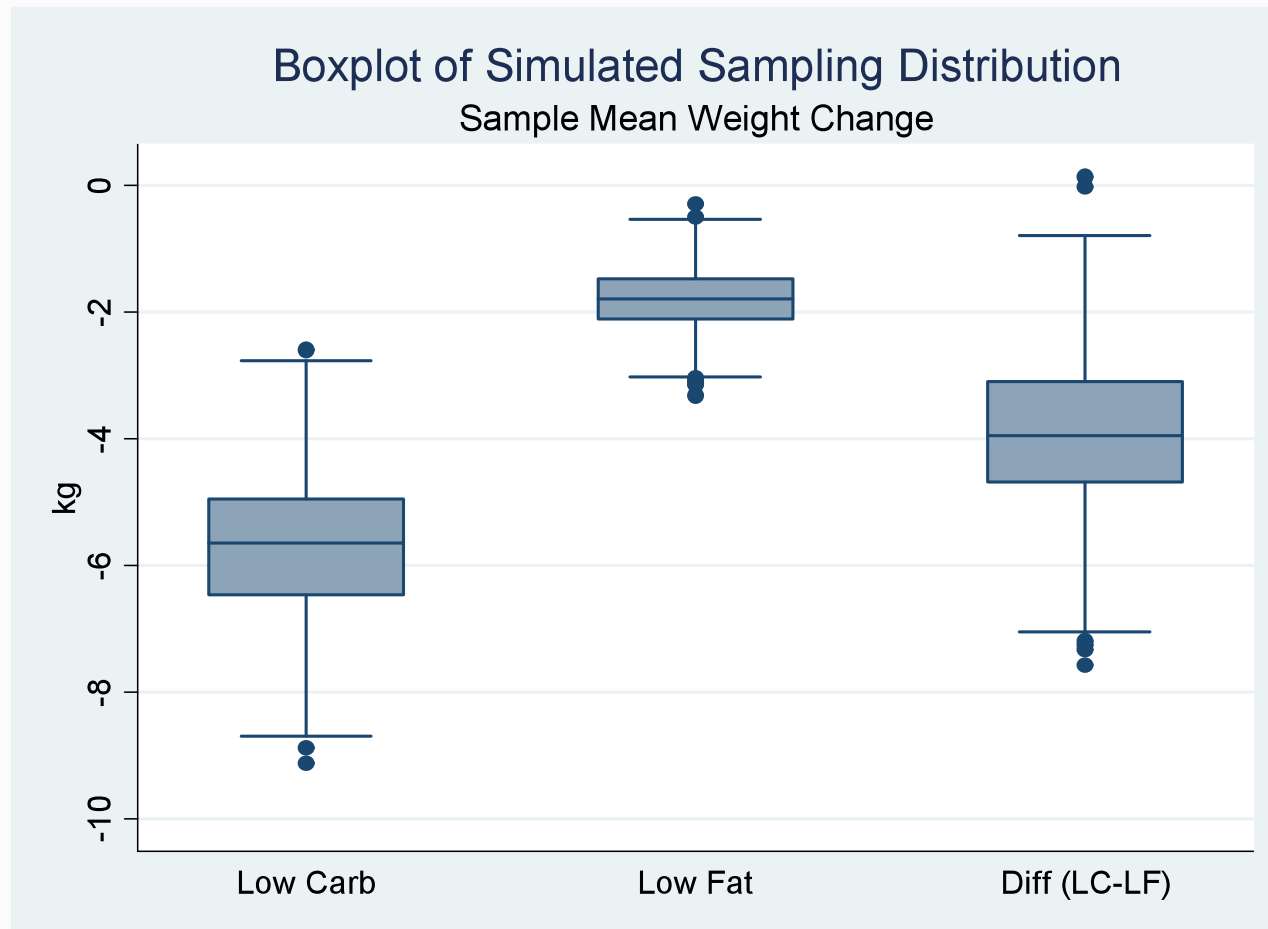
Simulated Sampling Dist'n of Sample Mean Weight Loss

- Simulated sampling distribution of sample mean weight change: low fat diet group



Simulated Sampling Dist'n of Sample Mean Weight Loss

- Side by side boxplots



95% Confidence Interval for Difference in Means

- Our most general formula

best estimate from sample $\pm 2 \times SE(\text{best estimate from sample})$

- The best estimate of a population mean difference based on sample means:

$$\bar{x}_1 - \bar{x}_2$$

- Here, \bar{x}_1 may represent the sample mean weight loss for the 64 subjects on the low carbohydrate diet, and \bar{x}_2 the mean weight loss for the 68 subjects on the low fat diet

95% CI for Difference in Means: Diet Types Study

- So, $\bar{x}_1 - \bar{x}_2 = -5.7 - (-1.8) = -3.9 \text{ kg}$: hence the formula for the 95% CI for $\mu_1 - \mu_2$ is:

$$-3.9 \pm 2 \times SE(\bar{x}_1 - \bar{x}_2)$$

- Where $SE(\bar{x}_1 - \bar{x}_2)$ = standard error of the difference of two sample means

Two Independent (Unpaired) Groups

- The standard error of the difference for two independent samples is calculated differently than we did for paired designs
 - With paired design we reduced data on two samples to one set differences between two groups
- Statisticians have developed formulas for the standard error of the difference
- These formulas depend on sample sizes in both groups and standard deviations in both groups
- The $SE(\bar{x}_1 - \bar{x}_2)$ is greater than either $SE(\bar{x}_1)$ or $SE(\bar{x}_2)$
 - Why do you think this is?

Principle

- Variation from independent sources can be added
 - Why do you think this is additive

$$SE(\bar{x}_1 - \bar{x}_2) = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

- Of course, we don't know σ_1 and σ_2 : so we estimate with s_1 and s_2 to get an estimated standard error:

$$S\hat{E}(\bar{x}_1 - \bar{x}_2) = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} = \sqrt{S\hat{E}(\bar{x}_1)^2 + S\hat{E}(\bar{x}_2)^2}$$

Comparing Two Independent Groups: Diet Types Study

- Recall the data from the weight change/diet type study

	Diet Group	
	Low-Carb	Low-Fat
Number of subjects (n)	64	68
Mean weight change (kg) Post-diet less pre-diet	-5.7	-1.8
Standard deviation of weight changes (kg)	8.6	3.9

$$SE(\bar{X}_1 - \bar{X}_2) = \sqrt{\frac{8.6^2}{64} + \frac{3.9^2}{68}} \approx 1.17$$

95% CI for Difference in Means: Diet Types Study

- So in this example, the estimated 95% for the true mean difference in weight between the low-carbohydrate and low-fat diet groups is:

$$-3.9 \pm 2 \times S\hat{E}(\bar{x}_1 - \bar{x}_2)$$

$$-3.9 \pm 2 \times 1.17$$

$$-3.9 \pm 2 \times 1.17$$

$$-6.24 \text{ kg to } -1.56 \text{ kg} \approx$$

$$-6.2 \text{ kg to } -1.6 \text{ kg}$$

From Article

- “Subjects on the low-carbohydrate diet lost more weight than those on a low fat diet (95% confidence interval for the difference in weight loss between groups, -1.6 to -6.2 kg; $p < .01$)”
- So those on the low carb diet lost more on average by 3.9 kg: after accounting for sampling variability this excess average loss over the low-fat diet group could be as small as 1.6 kg or as large as 6.2 kg
 - This confidence interval does not include 0, suggesting a real population level association between type of diet (low-carb or low-fat) and weight loss